

如何确保生成式医学人工智能发展可控

●本报记者 张思玮 见习记者 张帆

近日,由广州市卫生健康委员会主办的“广州人工智能与传染病防治学术交流会”在国家医学人工智能应用示范中试基地(广州)举办。会上正式发布了《生成式医学人工智能(GMAI)临床应用伦理治理专家共识(2025)》(以下简称《共识》)。

“近年来,GMAI发展迅猛,也越来越多地被用在临床一线,正在深刻改变我们提供医疗服务的方式和内涵。这不仅是技术革命,更是一场深刻的医疗伦理与治理革命。”北京协和医院科研处教授白桦表示。

“GMAI在辅助诊断、制定个性化治疗方案、改善医患沟通,以及推动科研创新等方面,都展现出巨大的潜力,给提升医疗质量和优化资源配置带来了前所未有的机会。”中山大学附属第一医院副院长陈巍介绍说,“但技术跑得太快,也带来不少问题,比如患者的数据隐私怎么保护,算法会不会出现偏倚,出了问题谁来负责,临床应用中的风险又该怎么控制。GMAI所特有的‘生成幻觉’、算法偏倚与责任模糊等风险不容忽视,而当前的伦理治理明显滞后于技术落地。”

通过明确阈值 保障 AI 决策可靠性

据了解,《共识》的主体内容由“核心治理风险”“技术适配性与公平性保障”“全生命周期监管与责任机制”和“社会文化嵌入与医患关系维护”四大板块组成,并最终形成29条主题化的“推荐意见”形式呈现,涵盖产品从技术安全到人文关怀的多个维度。

《共识》提出“生成幻觉阈值控制”“置信度标注”“多模态交叉验证”等技术治理要求。GMAI在输出诊疗方案或处置建议时,应提供量化置信度评分(如0~100分),或置信强度等级,直观反映推理依据与可靠程度,帮助医生快速判断结果可信度并决定是否进一步验证或会诊。此外,应同步建立冲突预警机制,实时比对患者既往病史、实验室数据及医学常识。一旦发现AI建议与实际情况存

“期待《共识》能够持续回应临床实践与技术发展的最新挑战,保障GMAI在安全、合规和伦理可控的轨道上稳健发展”。

在矛盾,应立即触发提示或强制复核,从源头降低模型幻觉、数据偏差等引发的风险。

“所谓‘生成幻觉’,就是指GMAI在给出诊断建议、解释医学文本时,会输出缺乏事实依据、与真实医学信息不符或完全虚构的信息。这是GMAI特有的风险,可能是由训练数据偏差、模型过度泛化或者推理错误导致的。”广东医科大学多模态数据融合创新实验室(GMC Lab)主任、《共识》执笔组组长弓孟春介绍。

基于此,《共识》要求,应依据场景风险分级设定可量化的幻觉率上限,在初诊、手术方案制定等直接关系到生命的高风险环节,幻觉率应低于3%;而在慢病复诊或健康管理等相对低风险场景,则建议放宽至低于5%。同时,幻觉率的阈值控制区间应随技术发展进行动态调整。此外,应推行多模态交叉验证,要求GMAI将文本诊断与影像、检验等多源数据自动比对,发现矛盾即时预警,以降低“生成幻觉”的发生率和危害。

为实现AI与临床实践稳健融合、确保技术进步真正服务于高质量医疗,《共识》还提出建立GMAI决策链可视化溯源机制、GMAI应用紧急熔断机制、GMAI输出的重大诊疗建议(如医生应进行最终审核与签字确认)。

让 GMAI 真正实现 普惠化与均衡化

据了解,基层医疗机构常常面临服务器性能不足、网络带宽有限等硬件约束,导致大型GMAI应用难以顺畅运行。

AI如何真正落地基层医疗体系?《共识》指出,应通过模型压缩、参数剪枝与知识蒸馏等技术,将模型参数

量进行一定的控制,同时保持诊断精度和核心功能,从而在更低计算资源下实现稳定部署。此外,应针对社区医院、乡镇卫生院等基层单位的医生开展系统化培训,重点包括GMAI的基本原理、数据输入与输出流程、常见故障处理以及隐私保护与伦理规范。培训内容需结合基层常见病、多发病的实际案例,并提供模拟操作,让医生能够熟练完成AI辅助问诊、诊断和报告解读。

《共识》高度重视群体公平,关注方言群体、文化程度较低者及特殊人群等,建议优化GMAI功能,以支持至少七大方言的语音与文本交互,并提供容错机制,体现出对多元化国情与医疗资源不平衡现实的深刻回应。

“比如,当系统听到‘咳嗽’这个词,但因为患者说方言导致发音不太清楚时,就可以自动弹出提示询问,这样患者就能马上确认或纠正。”厦门大学医学院生命伦理学中心主任、《共识》执笔组专家成员之一马永慧举例说,“这样做不仅让GMAI在各种方言环境下更好用,也让老年人和基层用户觉得更贴心、更值得信赖。通过优化多方言交互,GMAI就能真正走进更多本土的医疗场景,让智能医疗实实在在地惠及更多人。”

患者有权向 AI 说“不”, AI 不替代人文

相关调研结果显示,当医生使用AI问诊系统时,平均会有63%的时间注视屏幕,这导致患者感知到的共情水平下降了41%。

GMAI是否会改变传统的医患互动模式?患者是否有权拒绝使用AI?

《共识》规定,在GMAI辅助诊疗场景中,医生必须保留充足的面对面交流时间,根据诊疗复杂度设定最低标准,如初诊不少于15分钟、复诊不少于10分钟。医生需在完成AI辅助诊断后,向患者详细解释诊疗结果与治疗方案,并确认患者理解与同意,避免“机器替代沟通”造成的信任缺失或信息遗漏。

《共识》也强调,医生在使用

GMAI前,应在知情同意书中明确告知患者的选择权,包括拒绝AI参与诊疗的权利。即便患者拒绝使用AI辅助,医生与医疗机构也必须提供与传统方式同等质量的医疗服务、同等标准的传统诊疗方案,不得因患者选择而降低服务水平或延误治疗,从而保障患者的知情同意和自由决定权。

需明确医生为 诊疗决策最终责任人

《共识》的主要适用对象涵盖医疗机构管理层、临床一线医生、医疗AI研发企业以及监管与伦理治理部门。

值得注意的是,《共识》明确,临床一线医生是直接操作和评估GMAI输出的关键群体,需要掌握AI工具的判读、复核与患者沟通技能。AI仅作为辅助工具,医生须结合患者具体情况对AI建议进行审核与判断,医生始终是诊疗决策的最终责任人,对采纳AI建议后的临床结果负有不可推卸的伦理与法律责任。这既避免了AI成为“责任黑洞”,也防止医生因过度依赖技术而导致技能退化。

此外,医疗机构管理层负责在医院内部建立完善的GMAI使用规范、质量控制与风险预警体系,并对系统部署环境适宜性及流程管理所引发的运营性风险承担伦理责任;医疗AI研发与运营企业对模型算法偏差、数据缺陷及由此产生的“生成幻觉”承担源头治理责任,在产品迭代中全面落实《共识》的技术与伦理要求;监管与伦理治理部门则负责制定配套政策、监督执行情况并定期开展审计,特别是需肩负起防范与化解全局性公共风险的伦理治理责任,确保《共识》的规范实施。

“为保障前瞻性与实用性,建议对《共识》每两年开展一次系统化更新。”中华医学会医学伦理学分会副主任委员、南方医科大学南方医院赣州医院党委书记戴辉表示,通过动态反馈与迭代机制,“期待《共识》能够持续回应临床实践与技术发展的最新挑战,保障GMAI在安全、合规和伦理可控的轨道上稳健发展”。